

A formal account of the interaction of orthography and perception

English intervocalic consonants borrowed into Italian

Silke Hamann¹  · Ilaria E. Colombo¹

Received: 3 September 2015 / Accepted: 25 January 2017 / Published online: 27 March 2017

© The Author(s) 2017. This article is published with open access at Springerlink.com

Abstract This study presents a formal generative model that integrates perception and reading, and uses English intervocalic consonants borrowed into Italian as either singletons or geminates to illustrate how the model works. Consisting of words borrowed in the 20th century, our data show that the quantity of the intervocalic consonant in an Italian loanword depends on its written representation in English, the source language. Thus only English intervocalic consonants that are written with two identical letters (for example, as in *splatter*) are borrowed as geminates. We provide a formalization of these orthographic adaptations with grapheme-to-phoneme mappings in the shape of Optimality-theoretic constraints that model the native reading process, and show how the output of these mappings is restricted by native phonotactic constraints. Furthermore, we illustrate that the native reading grammar proposed here complements the perceptual adaptation model by Boersma and Hamann (2009). This combined model is shown to be able to account for simultaneous orthographic and perceptual borrowings in Italian, as well as to hold for reading and perception outside the realm of loanword adaptation.

Keywords Phonology · Loanwords · Orthography · Perception · Reading grammar · Italian

1 Introduction

Several studies on loanword adaptations state the important role orthography can play in the adaptation process. Friesner (2009), for instance, illustrates that Romanian

✉ S. Hamann
silke.hamann@uva.nl

I.E. Colombo
i.e.colombo@uva.nl

¹ Universiteit van Amsterdam, Amsterdam, The Netherlands

loans of French words with “final orthographic consonants that are not pronounced in French are occasionally realized in Romanian loans” (128), e.g. the French word *boulevard* [bul(ə)var], which is borrowed as [bulevard] in Romanian. Smith (2006) found that in Japanese, loan doublets (borrowings of the same words twice with different resulting forms) often stem from separate means of borrowing, one perceptual and the other orthographic. Further examples of real adaptations that show an influence of orthography are given e.g. by Kang (2009) for Korean and Miao (2005) for Mandarin. A number of experimental studies on second-language perception, which are often interpreted as imitations of online loanword adaptations, show that writing can have a positive influence on the correct perception and identification of L2 segments; see e.g. the studies by Vendelin and Peperkamp (2006), Detey and Nespoulous (2008), Escudero et al. (2008), Escudero and Wanrooij (2010), Porter (2010), and Daland et al. (2015) and the overview by Bassetti et al. (2015). What the literature lacks, however, is a formalization of such an orthographical influence; that is, how the written form must be incorporated into a formal grammar model to account for the observed effects.

In this article we provide a model that accounts for orthographic borrowings by analyzing loanword adaptations in Italian, a language with a relatively transparent grapheme-to-phoneme mapping. Following Coltheart et al. (1993), we use ‘grapheme’ to refer to any letter or group of letters that corresponds to a single phoneme. Italian has a singleton-geminate contrast in consonantal length in intersonorant position, which is reflected in the orthographic representation of the consonants. In loan adaptations into Italian, intervocalic consonants in words from languages that only have singletons, such as English, are often borrowed as geminates, e.g. *hobby* /ˈɒbbi/. Although accounts of Italian loanword adaptations abound, e.g. Rando (1970), Repetti (1993, 2009, 2012), Morandini (2007) and Passino (2008, 2013)—and most do mention an influence of orthography on the adaptation of consonants—none provides a grammar model that can account for this influence.

The present study shows that for Italian borrowings from English in the 20th century the quantity of the intervocalic consonant in the Italian loanword depends on its written representation in English. More precisely, only English intervocalic consonants that are written with two identical letters are borrowed as geminates. To account for such an orthographic influence on the borrowing process, we provide the formalization of a native reading grammar that maps a written form onto a phonological surface form, where the output is restricted by native phonotactic constraints. This native reading grammar alone is shown to be able to account for the attested orthographic effects.

In the following section of this article, we introduce Italian syllable and word structure. Section 3 provides the loanword data, illustrates the influence of orthography on the borrowing of intervocalic consonants, and shows an interaction of orthography and perception in the borrowing process for some of these words. Section 4 formalizes the native reading grammar and its possible interaction with speech perception in Optimality Theory, henceforth ‘OT’ (Prince and Smolensky 1993 [2004]). Section 5 shows briefly what a reading grammar looks like for two identical consonant letters in German native and non-native words; further, it discusses the implications of the reading grammar for a larger grammar model of linguistic knowledge,

the possible modelling of the writing process, and the possible modelling of reading of languages with a less transparent grapheme-phoneme mapping than Italian. Section 6 compares the present proposal to earlier formal accounts of reading, and to earlier accounts of orthographic loanword adaptation in Italian. In Sect. 7, we offer some conclusions.

Before introducing Italian phonology, a remark on the employed notation is in order. In this article we use pipes for underlying, lexical representations; slashes for surface, allophonic representations; square brackets for auditory forms; and angle brackets for written forms. Geminates in surface phonological representations are transcribed by two separate identical consonant symbols (rather than one symbol with an additional length sign) as this allows a positioning of the syllable boundary between the two. Auditory forms stand for concrete values along continuous auditory dimensions such as first formant, second formant, duration, etc. They are given in IPA transcriptions though should not be confused with abstract phonological categories such as allophones or phonemes which are also transcribed with IPA symbols, albeit in either slashes or pipes.

2 Italian phonology in brief

This section looks at the phoneme inventory of Standard Italian and then moves on to phonotactic restrictions. An overview of the Italian singleton consonants is given in (1) (based on Bertinetto and Loporcaro 2005).

(1) *Italian singleton consonants*

	Labial		Alveolar		Postalveolar	Palatal	Velar
Plosives	p	b	t	d			k g
Fricatives	f	v	s	z	ʃ (ʒ) ¹		
Affricates			ts	dʒ	tʃ dʒ		
Nasals		m		n		ɲ	
Laterals				l		ʎ	
Glides						j	w
Trills				r			

Most Italian consonants have a length contrast (singleton vs. geminate) in word-internal position when a vowel precedes and a vowel, glide or liquid follows. Exceptions are | z j w ʒ |², which have no geminate counterparts intervocally, and | ɲ: ʃ: ʎ: tʃ: dʒ: |, which can only occur geminated in this position and are therefore sometimes called intrinsically or inherently long (e.g. Passino 2008 or Repetti 2009).

The vowel system of Italian consists of seven phonemes | a e ε i o ɔ u |, all of which can occur in stressed position. In unstressed position, the lax vowels | ε ɔ | are prohibited. We assume in the following analysis that consonantal length contrasts are

¹/ʒ/ only occurs in loanwords, e.g. *beige* or *garage*.

²We follow Rogers and d'Arcangeli (2004), Krämer (2009), and others in treating |j w| as singletons that cannot geminate intervocally, but for a different approach see e.g. Marotta (1988).

stored underlyingly and that vowel length contrasts are not (see e.g. Repetti 1993 or Krämer 2009; for an alternative proposal see e.g. Saltarelli 1984). We further assume that geminates are parsed as heterosyllabic (see Saltarelli 1983; Loporcaro 1990).

Vowels show an allophonic length contrast in the surface representation: a stressed vowel in an open syllable is long (unless word final) but in a closed syllable it has to be short. We attribute this distribution to a bimoraic requirement on the stressed syllable; see the OT constraint in (2a) (cf. rule 2 by Repetti 1993:183; but for a formulation not referring to the mora, see e.g. Vogel 1982).³ The restriction on word final vowels to be short is formalized in (2b). In Italian, this constraint is ranked above constraint (2a) to ensure that final stress will not result in vowel lengthening.

(2) *Structural constraints relevant for the present account of Italian*

- a) /.'µµ./: Assign a violation mark to every stressed syllable that is not bimoraic.
- b) */V:#/ : Assign a violation mark to every word-final long vowel.
- c) */V:ʃV/ : Assign a violation mark to every intervocalic singleton /ʃ/.
- d) */VzV/ : Assign a violation mark to every intervocalic geminate /z/.
- e) PENULT: Assign a violation mark to every word with non-penultimate stress.

Furthermore, we employ constraints such as (2c) and (2d) to formalize (some of) the idiosyncratic restrictions on intervocalic singletons and geminates in Italian mentioned above.

Stress in non-derived nouns is most frequently on the penultimate syllable (for references, see Krämer 2009:161). We employ the constraint PENULT in (2e) to assign this default stress (following Repetti's 1993 rule 1).⁴ There are numerous exceptions, e.g. /.'ʃit.ta./ 'city' or /.'pɛ:ko.ra./ 'sheep', for which stress is assumed to be stored lexically.⁵

The constraints in (2) are sufficient for the analysis of singletons and geminates in native and borrowed words in the present article but do not cover all of Italian phonology; for a complete picture, see e.g. Krämer (2009).

3 The data

The data analyzed in the present study all stem from the dictionary by Zingarelli et al. (2015). For accuracy, they were checked with six native speakers: three older (average age of 63) and three younger (including this article's second author; average age of

³We disregard lengthening of the penultimate vowel in the present article, which applies to both open and closed syllables (see measurements by D'Imperio and Rosenthal 1999:6) and adds extra length to a stressed penultimate vowel.

⁴For alternative accounts that involve the assignment of a non-final trochaic foot via constraints, see D'Imperio and Rosenthal (1999) and Krämer (2009).

⁵A faithfulness constraint referring to lexical stress would have to be included in the analysis and ranked above PENULT to account for such irregular stress patterns.

25). The native speakers largely agreed with the pronunciations given in Zingarelli; for a full list of their responses, see the [Appendix](#).⁶

We focus on Italian loanwords borrowed from English in the last century (which makes an orthographic influence in the borrowing process more likely) that have an intervocalic consonant preceded by a short/lax vowel in English, as this can result in either a singleton or a geminate in the Italian loanword. Preceding long/tense vowels in the English form result in a borrowing with a singleton, see e.g. *slogan* /.'zlɔ:.gan./ or *speaker* /.'spi:.ker./ for two reasons. First, a vowel with long duration in relation to a shorter following consonant is perceptually interpreted by Italian native speakers as a vowel followed by a singleton (Esposito and Di Benedetto 1999; Pickett et al. 1999). Second, long/tense interconsonantal vowels in English are always written with only one following consonantal letter, and this single consonantal letter does not cause an orthographic interpretation as geminate. Orthographic and perceptual information therefore would result in the same adaptation of a consonant preceded by a long/tense vowel, namely as singleton.

We base our decision whether the preceding vowel is tense or lax on a Standard Southern British English variety, because we assume that this is the variety that native speakers living in Italy are more exposed to (or at least were in the first half of the 20th century). Where relevant, we discuss possible alternatives. For an account of the incorporation of American English words into the Italian of Italian immigrants, see e.g. Repetti (2009).

The dataset discussed in this section is an exhaustive list of all loanwords given in Zingarelli et al. (2015) that were borrowed from English in the 20th century and have a short/lax vowel preceding an intervocalic consonant (based on a British English pronunciation). In (3) are all words that were adapted with an intervocalic geminate: the words in (3a) have stress on the vowel preceding the consonant in English, and those in (3b) have stress on the following vowel. For the examples given in this section, the date after each word refers to its first attestation (as given in Zingarelli et al.).⁷ The relevant consonant(s) are given in boldface.

(3) *Borrowings with geminate consonants*

a)	/.'ban.ner./	<i>banner</i>	1996
	/.'ɒb.bi./	<i>hobby</i>	1952
	/.'ɔr.ror./	<i>horror</i>	1977
	/.'dʒɒl.li./	<i>jolly</i>	1923
	/.'ip.pi./	<i>hippie</i>	1967
	/.'ʃɒp.pin./	<i>shopping</i>	1931
	/.'tril.ler./	<i>thriller</i>	1957

⁶We chose the dictionary by Zingarelli et al. (2015) as our source because it is updated frequently, contains many loanwords, and seems to reflect the actual pronunciation of these words quite well, as the comparison with the production of our six native speakers shows.

⁷We adjusted Zingarelli et al.'s transcription by indicating vowel length in stressed syllables, and by changing schwas that Zingarelli et al. employed for some unstressed vowels into /e/, in accordance with the intuition of our native speakers.

English words ending in *-ing* show variation in the final sounds of the borrowed form between /-in/ and /-ing/ (with nasal place assimilation) (Zingarelli et al. 2015), where younger speakers seem to exclusively use the latter (as confirmed by our younger speakers). This variation is not included in our transcriptions.

	/.'rɛl.li./	<i>rally</i>	1935
	/.'baf.fer./	<i>buffer</i>	1983
	/.'splat.ter./	<i>splatter</i>	1986
	/.'dʒɔŋ.gin./	<i>jogging</i>	1978
	/.'njuz.'lɛt.ter./	<i>newsletter</i>	1985
	/.'ɛp.pe.nin./ ~ /.'ap.pe.nin./	<i>happening</i>	1964
	/.'be.bi.'sit.ter./	<i>baby sitter</i>	1950
	/.'no.'kɔm.mɛnt./	<i>no comment</i>	1963
b)	/.'ak.'kaunt./	<i>account</i>	1987
	/.'at.'tɛtʃ.mɛnt./	<i>attachment</i>	1994
	/.'kɔm.'man.do./	<i>commando</i>	1900
	/.'pul.'lɔ:.ver./	<i>pullover</i>	1927

The loanwords in (4), on the other hand, which also have an intervocalic consonant preceded by a short/lax vowel in the English words, were borrowed into Italian with a singleton consonant.⁸ Importantly, the preceding vowels (which are short in English) in all these words are borrowed as long.

(4) *Borrowings with singleton consonants*

	/.'ɔ:.kei./	<i>hockey</i>	1927
	/.'a:.ker./	<i>hacker</i>	1986
	/.'ɛ:.di.tor./	<i>editor</i>	1962
	/.'glɛ:.mur./	<i>glamour</i>	1953
	/.'ka:.me.ra.men./	<i>cameraman</i>	1959
	/.'mɔ:.ni.tor./	<i>monitor</i>	1963

The different borrowing strategies in (3) and (4) cannot be attributed to differences in quality or quantity of the vowel, duration, place or manner of articulation of the consonant,⁹ or stress pattern of the English word forms. We can therefore exclude an explanation in terms of specific perceptual cues that lead to an adaptation as singleton or geminate. Both sets of words were borrowed in the last century, with a similar spread across this time span, thus a diachronic change in adaption strategy (as e.g. proposed by Hamann and Li 2016 for borrowings into Hong Kong Cantonese) also has to be excluded as explanation for the present data. Phonotactic restrictions cannot account for the two adaptation patterns, either, because phonotactically, both singleton and geminate are possible, e.g. /.'ban.ner./ and /.'ba:.ner./ for *banner* or /.'ɔk.kei./ and /.'ɔ:.kei./ for *hockey*; see the existing adaptations of near-minimal pairs such as *hobby* (with singleton) vs. *hockey* (with geminate), or *glamour* (with singleton) vs. *banner* (with geminate).

Orthography, on the other hand, is clearly correlated with the choice of consonantal length: all the consonants borrowed as geminates in (3) are written with two iden-

⁸The words in (4) all have stress on the first syllable. Loanwords in Zingarelli et al. (2015) with singletons and different stress patterns are all (nominalizations of) phrasal verbs. We come back to such cases in Sect. 6.1 when discussing the study by Repetti (1993).

⁹Some consonants occur only in one of the two sets (e.g. /k/ as singleton or /p/ as geminate). We put this down to accidental gaps in the data sets and refrain from postulating idiosyncratic borrowing strategies such as “/Vkv/ is always adapted as /V:kV/.”

tical letters, while the consonants borrowed as singletons in (4) are written either with a single letter or two different letters (<ck>). Italian geminates are orthographically represented by two identical letters (or two identical letters followed by another letter, e.g. <cc(i)> for /tʃ/). The only exceptions are the intrinsically long consonants /ɲ/ ʃ/ ʌ/ tʃ/ dʒ/, which are written with single consonant letters (though /tʃ/ dʒ/ can also be written as <zz>), and the intervocalic sequence <cqu> which is /k.kw/ (e.g. *acqua* /'ak.kwa./). Singleton consonants are usually written with a single letter. Exceptions to this one letter-singleton generalization are <ch> for /k/ and <gi> for /dʒ/ before non-front vowels, and <gn> for /ɲ/, <gl(i)> for /ʎ/, and <sc(i)> for /ʃ/ word-initially; for further details see e.g. Hall (1944). We therefore propose that the borrowers apply their knowledge of Italian grapheme-to-phoneme correspondences to the English written form when adapting these words.

In our dataset there are three exceptions to this proposed orthographically-based adaptation strategy for intervocalic consonants, given in (5).

(5) *Borrowings that violate the orthographic prediction*

/.'fɛʃ.fon./	<i>fashion</i>	1905
/.'pa:.zel./	<i>puzzle</i>	1919
/.'mɔ:bin./	<i>mobbing</i>	1992

In the first instance in (5), the consonant is borrowed as geminate although written with two different letters (<sh>), in the second it is borrowed as singleton though written with two identical letters (<zz>). Both exceptions are due to idiosyncratic phonotactic restrictions of Italian: in intervocalic position, /ʃ/ only occurs as geminate (recall restriction (2c)), and /z/ only as singleton (recall restriction (2d)). We therefore assume that native phonotactic restrictions restrain the output of orthography. This phonotactic influence can also be observed in the fact that borrowed forms with a short vowel followed by a short prevocalic consonant are not allowed, e.g. */.'ba.ner./ or */.'pa.zel./: the phonotactic restriction that stressed vowels have to be bimoraic, recall constraint (2a), is prohibiting such outputs. This interaction of orthographic mappings with phonotactic constraints will be formalized in Sect. 4.1 below.

The third exception in (5), /.'mɔ:bin./, is not due to phonotactic restrictions, as a form with a geminate (/.'mɔb.bin./) is not only possible but also the only one that our six native speakers used. According to these native speakers' judgments, *mobbing* therefore forms no exception and is borrowed with a geminate, as the English orthography would predict.

For the words in (3) and (4) above we argued that the orthography determines the quantity of the consonant (and the correlating quantity of the preceding vowel). A further indicator for an orthographic influence in the adaptation of these words is the quality of the vowels preceding the singletons/geminates. Several of them reflect the Italian grapheme-phoneme mapping for vowels instead of the English pronunciation, see e.g. *banner* /.'ban.ner./, *splatter* /.'splat.ter./ and *hacker* /.'a.ker./. In these cases the English [æ] has not been rendered with the perceptually closest Italian vowel /ɛ/ (see e.g. the results of the perception experiment by Flege et al. 1999), but with /a/, the vowel corresponding to the grapheme <a> in Italian. For words like these, we can therefore assume a purely orthographic borrowing process, which we formalize in Sect. 4.2 below.

The borrowed vowels in other words are clearly influenced by the perception of the English form, as e.g. the stressed vowels in *buffer* /.'baf.fer./, *glamour* /.'glɛ:.mur./ and *rally* /.'rɛl.li./. For cases like these, we have to assume an interaction of perceptual and orthographic borrowing strategies, where perceptual cues to vowel quality and orthographic mappings together determine the output (again restricted by native phonotactic constraints). This interaction of orthographic and perceptual mappings is formalized in Sect. 4.3 below.

4 Modelling orthographic and perceptual borrowings

In this section, we formalize the orthographic adaptation of English intervocalic consonants into Italian by introducing an Optimality-Theoretic *reading grammar*, i.e. the language-specific mapping of written forms onto surface phonological forms, which is used in the reading process.

The working of an Italian reading grammar is illustrated with three native Italian words in Sect. 4.1. In Sect. 4.2, we show how orthographic borrowings from English can be accounted for by employing this Italian reading grammar to English written forms. This is of course only possible because both native and source language employ a Roman alphabetic script. In Sect. 4.3, we illustrate that our reading grammar is compatible with a perception grammar as proposed by Boersma (2007) and applied by Boersma and Hamann (2009) to loan adaptations, and that the combined model can account for possible cases of simultaneous orthographic and perceptual borrowings.

4.1 Native reading: Orthographic mappings and phonotactic constraints

In the process of reading alphabetic scripts, a written form is mapped onto a phonological surface form (henceforth: SF). The latter is then used to retrieve meaning from the stored form-meaning pairs in the mental lexicon (where 'form' is the phonological underlying form). The mapping between grapheme and SF is formalized in the present article with what we call *orthographic constraints* (ORTH). These orthographic constraints have the form as in (6).¹⁰

(6) *General orthographic constraints for shallow orthographies*

- a) $\langle \gamma \rangle / P/$: Assign a violation mark to every grapheme $\langle \gamma \rangle$ that is not mapped onto the phonological form /P/ and vice versa.
- b) $*\langle \gamma \rangle / /$: Assign a violation mark to every grapheme $\langle \gamma \rangle$ that is mapped onto an empty segment in the SF.
- c) $*\langle \rangle / P/$: Assign a violation mark if the absence of a grapheme is mapped onto the phonological form /P/.

Constraint (6a) is a constraint that is violated when $\langle \gamma \rangle$ is mapped onto any other SF than /P/ or any other grapheme than $\langle \gamma \rangle$ is mapped onto /P/. The constraints (6b) and

¹⁰As graphemes are mapped onto surface phonological forms, the latter can be allophones. We discuss the implications of this in Sect. 5.3 below. In the following, we employ the term *grapheme-phoneme mappings* to include mappings from graphemes onto allophones.

(6c) together express the (violable) orthographic principle “one letter—one sound” proposed by Wiese (2004).

In principle, a universal set of such orthographic constraints can be assumed, mapping all possible written units (including an empty form) onto all possible phonological surface units (including an empty form). It seems more plausible to us, however, that the language learner postulates such constraints on the basis of the acquired phonological surface units and the encountered written units, as this drastically reduces the number of constraints the learner has to handle and reflects the fact that the acquisition of reading (and writing) depends on previously acquired phonological knowledge (further discussed on the next page). Under this assumption, the constraints in (6) could be considered templates that learners employ to create language-specific ORTH constraints.

Examples of ORTH constraints of the shape (6a) that are relevant (i.e. high-ranked) in Italian are e.g. $\langle f \rangle / f/$, $\langle t \rangle / t/$, $\langle u \rangle / u/$, $\langle a \rangle / a/$, etc., but also e.g. $\langle gn \rangle / \eta/$, $\langle gli \rangle / \lambda/$, i.e. constraints with graphemes that consist of two or more letters and therefore violate the “one letter—one sound” principle. Such constraints have to be higher ranked than the constraint against an “empty” mapping in (6b) and constraints mapping single letters onto phonemes, to ensure the retrieval of the correct SF in words such as e.g. $\langle gnocchi \rangle / .\eta \partial k.ki./$ or $\langle aglio \rangle / .a\lambda.\lambda o./$ ‘garlic’. Languages like Italian with so-called *shallow alphabetic writing systems* (Lieberman et al. 1980; Katz and Feldman 1983), where the spelling is consistent, have mostly a one-to-one relationship between grapheme and phoneme, and mostly graphemes that consist of single letters. Languages like French and English with so-called *deep alphabetic writing systems* have more graphemes that consist of several letters, very often several graphemes for the same phoneme, and the same graphemes for different phonemes.

To account for the singleton-geminate distinction and its orthographic representation in Italian, we need to make a distinction between graphemes referring to consonants, $\langle \beta \rangle$, and those referring to vowels, $\langle \alpha \rangle$. The constraint (7a) is necessary to ensure that only a grapheme of two identical consonantal letters is mapped onto geminates in SF, and that only geminates in SF are mapped onto such a grapheme.¹¹

(7) *Orthographic constraints relevant for the singleton-geminate contrast in Italian*

- a) $\langle \beta_i \beta_i \rangle / C:/$: Assign a violation mark if a grapheme of two identical consonantal letters is not mapped onto a surface geminate, and vice versa.
- b) $*\langle \alpha \rangle / V:/$: Assign a violation mark whenever a single vowel letter is mapped onto a long surface vowel.
- c) $\langle \grave{\alpha} \rangle / 'V/$: Assign a violation mark to every vocalic letter with a grave accent that is not mapped onto a stressed vowel.
- d) $\langle h \rangle / /$: Assign a violation mark to every letter $\langle h \rangle$ that is not mapped onto an empty segment in SF.

¹¹This constraint collapses two negatively formulated constraints, $*\langle \beta_i \beta_i \rangle / C/$ and $*\langle \beta \rangle / C:/$, that do not need to be distinguished in the following analyses.

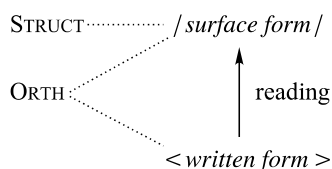


Fig. 1 Reading grammar: mapping of a written form onto a phonological surface form via ORTH(OGRAPHIC) constraints, and their interaction with STRUCT(URAL) restrictions that hold on the surface form

Constraint (7b) avoids that a single vowel letter is interpreted as a long vowel (with two moras), but since this happens quite often in Italian, namely every time a stressed vowel precedes a single intervocalic consonantal letter, this constraint is relatively low-ranked (and is not decisive in the following analyses).

The constraint in (7c) is included to illustrate how orthographic markings of non-default stress patterns are dealt with. In this case, the grave accent on the vowel letter has to be mapped onto a corresponding stressed vowel. Italian has further possibilities to mark irregular stress orthographically, which are not included in the present account. Constraint (7d) we employ to account for the fact that the letter <h> maps onto no SF in Italian.

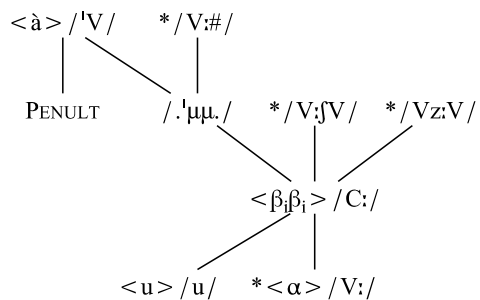
In Sect. 2 above we provided evidence from Italian that the output of the grapheme-phoneme mapping is influenced by phonotactic restrictions: this captures the idea that readers only create phonological forms that are in line with the phonological structure of their language. It furthermore prevents orthographic mappings from reduplicating phonological knowledge that is already represented somewhere else in the readers' grammar/brain. Neurolinguistic studies on reading alphabetic orthographies support our proposal: during the reading process, a cluster in the left inferior parietal gyrus is activated, which is usually also involved in non-reading related, sub-lexical phonological processes (see the meta-analysis of existing neuroimaging studies by Cattinelli et al. 2013). The assumed influence of phonological knowledge on the reading process furthermore predicts that a phonological deficit or difficulties in accessing phonological representations lead to problems in the acquisition of reading, which has been shown by studies on dyslexia (see e.g. Liberman and Shankweiler 1985; Ramus and Szenkovits 2008).

The proposed reading grammar thus looks as depicted in Fig. 1.

Section 5.3 below deals with possible orthographic mappings other than those onto surface phonological forms.

How are the ORTH constraints in (7) now ranked with respect to the STRUCT constraints introduced in (2)? The constraint < $\beta_i\beta_i$ >/C/ has to be dominated by the structural constraints /. $\mu\mu$./, */V: β V/ and */Vz:V/ to allow only phonotactically well-formed winners, which are the only attested forms, as we saw in the preceding section. Furthermore, the ORTH constraint < $\hat{\alpha}$ >/'V/ has to dominate the STRUCT constraints /. $\mu\mu$./ and PENULT to allow final non-bimoraic stressed vowels as output of the reading process. This will be illustrated in tableau (10) below. The ranking between STRUCT and ORTH constraints that we just established is depicted in Fig. 2, upper two rows.

Fig. 2 Italian reading grammar employed here



As for the lower part of Fig. 2, ORTH constraints of the form $\langle \gamma \rangle / P /$ referring to single letters, with $\langle u \rangle / u /$ as an example, are assumed to be lower ranked than the constraint $\langle \beta_i \beta_i \rangle / C: /$ because the mapping of two or more letters overrides single letter mappings. The ORTH constraint $* \langle \alpha \rangle / V: /$ is low ranked, too, as mentioned above, because it is violated often in the Italian reading process since the length of the vowels is not expressed in Italian orthography. The exact ranking of the ORTH constraint $\langle h \rangle / /$ cannot be determined on the basis of our data, as we will see in tableau (12) below, and the constraint is therefore not included in Fig. 2.

We postulated the ranking in Fig. 2 based on the frequency of occurring forms and on logical considerations. This ranking is learnable with the help of the Gradual Learning Algorithm (GLA; Boersma 1997; Boersma and Hayes 2001), assuming an initial ranking of all constraints at the same height, and gradual demotion and promotion on the basis of natural input distributions.

The following two tableaux of Italian illustrate the reading process for the native orthographic and phonological minimal pair $\langle \text{fatto} \rangle$ ‘done; fact’ and $\langle \text{fato} \rangle$ ‘fate’ with only the relevant constraints.¹²

(8) *Reading native <fatto>*

<fatto>	PENULT	/.'μμ./	<βiβi> /C:/	* <α> /V:/
fat.to.				
fa.to.			*!	*
fa.to.		*!	*	
fat.t.to.		*!		*
fat.to.	*!	*!		

The optimal candidate in this tableau is the first: the written form $\langle \text{fatto} \rangle$ is thus read as the surface form fat.to. . Candidates three, four and five all violate the high-ranked STRUCT constraint $/.'μμ./$, and candidate five additionally PENULT by having final stress. The second candidate satisfies all given structural constraints, but it violates $\langle \beta_i \beta_i \rangle / C: /$ because the double-letter grapheme is not mapped onto a long consonant.

In tableau (9), we formalize our assumption that the written form $\langle \text{fato} \rangle$ is mapped onto the phonological form fa.to. (the winning, second candidate), and not fa.to.

¹²Candidates with a geminate that do not span two syllables are not included in the present tableaux. They would violate an additional constraint requiring geminates to be bisyllabic.

(the third candidate), because this gives us a uniform bimoraic structure of the stressed non-final syllable in Italian. In order to attain this mapping, * $\langle\alpha\rangle/V:/$ has to be ranked lowest.

(9) *Reading native <fato>*

<fato>	PENULT	/.'μμ./	<β _i β _i >/C:/	*<α>/V:/
/.'fat.to./			*!	
☞ /.'fa:.to./				*
/.'fa.to./		*!		
/.'fa.'to./	*!	*!		

To illustrate the reading of a non-default stress pattern that is orthographically marked, we employ the native form <città> ‘city’ as input form in tableau (10). For this decision mechanism, the constraints <à>/'V/ and */V:#/ are relevant (and therefore included in the tableau), because without them the incorrect form /.'tʃit.ta./ would win, cf. the first candidate:

(10) *Reading native <città>*

<città>	*/V:#/	<à>/'V/	PENULT	/.'μμ./	<β _i β _i >/C:/	*<α>/V:/
/.'tʃit.ta./		*!				
/.'tʃi:.ta./		*!			*	*
☞ /.'tʃit.'ta./			*	*		
/.'tʃit.'ta:./	*!		*			*

The first and second candidate both violate the constraint <à>/'V/ referring to the orthographic stress mark, and show that this constraint has to be higher-ranked than the STRUCT constraint PENULT, which is violated by the winning, third candidate. Candidate four is in line with the written stress mark, but has a long, stressed final vowel, violating */V:#/ , which shows us that this constraint has to be higher ranked than the STRUCT constraint /.'μμ./.

The winning surface forms /.'fat.to./, /.'fa:.to./ and /.'tʃit.'ta./ from tableaux (8), (9) and (10) are mapped onto the underlying forms |fat:o|, |fato| and |tʃit:'a|, respectively. In this shape they are assumed to be stored together with their meaning in the mental lexicon of Italian speakers. The mappings of surface onto underlying form are not relevant for the present argument and therefore not formalized, but see Sect. 5.2 below for the full grammar model.

4.2 Orthographic borrowings: Reading non-native forms

The same Italian reading grammar, i.e. the structural and orthographic constraints and their ranking, that has been employed to formalize the native reading processes in Sect. 4.1 above, is able to account for orthographically borrowed forms, with the only difference that instead of native written forms, the input consists of English written forms. This is illustrated for English forms with double consonantal letters with the word *banner* in tableau (11).

(11) *Orthographic adaptation of <banner>*

<banner>	PENULT	/.'μμ./	<β _i β _i >/C:/	*<α>/V:/
[☞] /.'ban.ner./				
/.'ba.ner./		*!	*	
/.'ba:.ner./			*!	*
/.'ban.'ner./	*!			

The optimal candidate in the adaptation of <banner> via the native reading grammar is the first candidate, the attested form /.'ban.ner./. Candidates two and three, with a singleton consonant, violate the constraint that two identical consonantal letters should be mapped onto a geminate. Candidate four is unacceptable because it violates the STRUCT constraint PENULT.

For adaptations of English words containing two differing consonantal letters, as e.g. <hacker>, we need to account for the borrowing with a singleton instead of a geminate consonant. This is due again to the ORTH constraint <β_iβ_i>/C:/, which is violated when two differing consonantal letters are mapped onto a geminate, see the first candidate in tableau (12), compared to the second candidate. Candidate three violates the high-ranked /.'μμ./, and candidate four loses with respect to the winning second candidate because it has mapped initial <h> onto a phoneme /h/.¹³ As we can see in this tableau, the position of the constraint <h>/ / is of no relevance for the present evaluation.

(12) *Orthographic adaptation of <hacker>*

<hacker>	PENULT	/.'μμ./	<h>/ /	<β _i β _i >/C:/	*<α>/V:/
/.'ak.ker./				*!	
[☞] /.'a:.ker./					*
/.'a.ker./		*!			
/.'ha:.ker./			*!		*

In (5), we encountered two exceptions to the orthographic prediction on singleton-geminate adaptations: the word *fashion*, with an inherently long consonant that is written with two different consonantal graphemes, and the word *puzzle*, with a consonant that cannot be geminated in Italian but is written with two identical graphemes. We explained already that these exceptions can be captured via the native Italian phonotactic constraints */V:ʃV/ and */Vz:V/. Tableaux (13) and (14) provide the complete formalizations.

(13) *Orthographic adaptation of <fashion>*

<fashion>	* /V:ʃV/	PENULT	/.'μμ./	<β _i β _i >/C:/	*<α>/V:/
/.'fɛ:.ʃon./	*!				*
/.'fɛ.ʃon./			*!		
[☞] /.'fɛʃ.ʃon./				*	

¹³The form with an initial /h/ also violates a possible STRUCT constraint */h/ “Assign a violation mark to every surface /h/”, which would be ranked high as Italian does not have a glottal fricative in its phoneme inventory. This STRUCT constraint would attain the same result as the ORTH constraint <h>/ /.

(14) *Orthographic adaptation of <puzzle>*

<puzzle>	*/Vz:V/	PENULT	/.'μμ./	<β _i β _i >/C:/	*<α>/V:/
^{ep} /.'pa:zel./				*	*
/.'pa.zel./			*!	*	
/.'paz.zel./	*!				

These tableaux illustrate the necessity to rank the STRUCT constraints */V:ɸV/ and */Vz:V/ above the ORTH constraint <β_iβ_i>/C:/, as postulated in the full reading grammar in Fig. 2.¹⁴

4.3 Formalizing simultaneous orthographic and perceptual adaptations

The quality of the stressed vowel in the adaptation of the words *puzzle* /.'pa:zel./, *fashion* /'fɛʃ.fon./ and in several other loanwords in (3) and (4) is obviously not due to Italian grapheme-phoneme mappings. Instead, the English original vowel quality is mapped onto its auditorily closest Italian equivalent. Hence, [p^hʌzɪ] turns into Italian /.'pa:zel./, [fæʃən] into /'fɛʃ.fon./, [ɹæli] into /.'rɛl.li./, etc. In order to account for words like these, we have to model an interaction of orthographic influences (determining the quantity of the intervocalic consonant) and perceptual influences (determining the quality of the preceding vowel). For the latter, we assume Boersma's (2007) *perception grammar*, which was applied by Boersma and Hamann (2009) to account for perceptual adaptations of loanwords. In the perception grammar, an incoming auditory form is mapped onto a SF with the help of CUE constraints of the form [A]/a/: "map the auditory form [A] onto the phonological surface form /a/." The output of this perception is again influenced by phonotactic restrictions (the STRUCTURAL constraints) and their language-specific ranking. As is obvious from this description, the OT reading grammar proposed in the present article is parallel to Boersma's perception grammar. And similar to the presently proposed orthographic adaptation process via a native reading grammar, a perceptual adaptation is nothing else than perceiving a non-native auditory input with a native perception grammar (Boersma and Hamann 2009).

In Fig. 3, the reading grammar (left part) and perception grammar (right part) are depicted in one step; both together can model simultaneous reading and listening by an interaction of ORTH, CUE and STRUCT constraints.

To illustrate an integrated orthographic-perceptual adaptation, the two Italian loanwords *buffer* /.'baf.fer./ and *rally* /.'rɛl.li./ are taken as representative examples, where the quality of the stressed vowels is obviously determined by the English auditory forms: Purely orthographic adaptations in the two example words would have rendered /.'buf.fer./, due to the ORTH constraint <u>/u/, and /.'ral.li./, due to the constraint <a>/a/. These two ORTH constraints are thus overridden by perceptual information, i.e. in OT terms, they are outranked by CUE constraints.

¹⁴Purely orthographic adaptations show regular penultimate stress, as in the examples analyzed in the present section. Antepenultimate stress in loanwords like *happening* /'ɛp.pe.nin./ in (3), *monitor* /.'mɔ:ni.tor./ in (4), *karavan* /.'ka:ra.van./ or *musical* /.'mju:zi.kəl./ (the last two from Bafile 1999:210) seems to mirror the stress placement of the original English words (as proposed also by Bafile *ibid.*), and therefore suggests (at least partially) perceptual borrowing as described in Sect. 4.3 below.

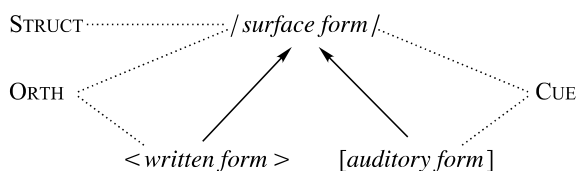


Fig. 3 The interaction of ORTH constraints (relevant for native reading and orthographic borrowings) with CUE constraints (for native perception and perceptual borrowings) in simultaneous orthographic and perceptual borrowings

The English vowel /ʌ/ in *buffer* is acoustically closest to the Italian vowel /a/ (Flege et al. 1999:2978) and also perceptually categorized as Italian /a/ by native Italian listeners (Flege and MacKay 2004:12). The same holds for English /æ/ in *rally* and its Italian equivalent /ɛ/ (ibid.). The two English–Italian vowel pairs /ʌ a/ and /æ ɛ/ are non-high, central to central-front vowels, which are similar acoustically and perceptually in that the first pair has mid second formant (F2) values, and the second mid to high F2 values. We therefore restrict the CUE constraints we employ here to determine their mapping to the auditory information of [mid F2] (for /ʌ/ and /a/) and [high F2] (for /æ/ and /ɛ/; as opposed to [very high F2] for more fronted vowels). These two cues (or rather ranges of values along the F2 dimension) are mapped onto the Italian native vowel categories /a/ and /ɛ/ with the high-ranked CUE constraints in (15a) and (15b), respectively.

- (15) *Cue constraints of Italian relevant for the integrated adaptations*
- a) [mid F2]/a/: Assign a violation mark to every auditory form with mid second formant values that is not mapped onto the surface vowel /a/.
 - b) [high F2]/ɛ/: Assign a violation mark to every auditory form with high second formant values that is not mapped onto the surface vowel /ɛ/.

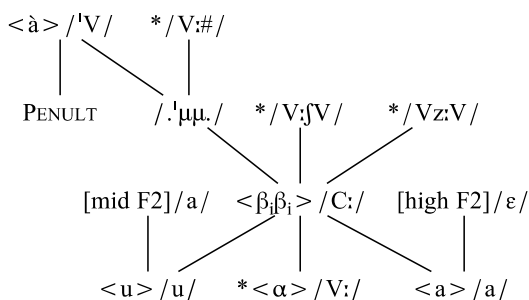
Further cues such as e.g. high amplitude formants that ensure the perception of a vowel, or first formant values to perceive the vowel height, are not considered.

We determined already that the CUE constraints (15a) and (15b) have to be ranked above the ORTH constraints <u>/u/ and <a>/a/, respectively, to predict the correct output forms for the integrated orthographic-perceptual account. This ranking expresses the fact that the perceptual cues for vowels are more salient than the orthographic forms of the vowels, if both percept and writing are available to the borrower.

These CUE constraints and their rankings together with the reading grammar we established at the end of Sect. 4.2 give us the reading and perception grammar in Fig. 4.

With this combined grammar, we can account for the Italian adaptations of *buffer* and *rally*, cf. tableaux (16) and (17). For lack of space, both tableaux only deal with the adaptation of the first syllable, and neglect stress assignment. Input forms are now both written and auditory forms. The auditory form gives only information on the second formant of the first vowel.

Fig. 4 Italian reading and perception grammar employed here



(16) *Combined orthographic and auditory adaptation of buffer*

<buffer> [Λ: mid F2]	/. 'μμ./	<βiβi>/C:/	[mid F2] /a/	[high F2] /ε/	<a>/a/	<u>/u/
/. 'ba:.fer./		*!				*
/. 'ba.fer./	*!	*				*
☞ /. 'baf.fer./						*
/. 'buf.fer./			*!			
/. 'bef.fer./			*!			*

The first two candidates violate the ORTH constraint against mapping two identical consonantal letters onto a geminate consonant. Candidate two shows an additional violation of the high-ranked STRUCT constraint requiring stressed syllables to be bimoraic. Between the structurally well-formed candidates three, four and five, the high-ranked CUE constraint [mid F2]/a/ decides: the input [Λ] is closest to Italian /a/ and therefore is perceived as such, even though the winning candidate, candidate three, violates the low-ranked ORTH constraint requiring the grapheme <u> to be mapped onto the phonological form /u/.

Exactly the same reasoning applies to tableau (17), where the winning, third candidate maps the auditory form with a high F2 onto /ε/, which is perceptually closer than /a/ (candidate four) or /e/ (candidate five). The winner violates a low-ranked ORTH constraint that requires a mapping of the grapheme <a> onto the surface form /a/.

(17) *Combined orthographic and auditory adaptation of rally*

<rally> [æ: high F2]	/. 'μμ./	<βiβi>/C:/	[mid F2] /a/	[high F2] /ε/	<a>/a/	<u>/u/
/. 'rɛ:.li./		*!			*	
/. 'rɛ.li./	*!	*			*	
☞ /. 'rɛl.li./					*	
/. 'ral.li./				*!		
/. 'rel.li./				*!	*	

In the two loanwords discussed in this section, we only looked at the perception of the vowel quality. But in forms like *puzzle* /.'pa:zel./, it is clearly not only the vowel quality that is determined by perception, but also the type of intervocalic consonant (native orthographic <zz> maps onto the affricates /d:z/ and /t:s/) and the order of segments in the final syllable, as alternative borrowings such as /.'pud.dzel/ and

/.'pud.dzle./ show (by one of our older speakers, cf. [Appendix](#)), the latter being fully orthographically borrowed. Purely perceptual borrowings can be modelled with a perception grammar, only, and the respective native CUE constraints. An illustration thereof would go beyond the scope of the present paper, but the interested reader is referred to Boersma and Hamann (2009) for the formalization of purely perceptual borrowings from English into Korean.

5 The bigger picture: Reading, comprehension and writing in Italian and other languages

Up to now we showed that a formal modelling of the native Italian reading process with an Italian reading grammar, where ORTH constraints interact with STRUCT constraints (Sect. 4.1), can also account for the borrowing of intervocalic consonants in loanwords from English (Sect. 4.2). We furthermore illustrated that this reading grammar can interact with the native Italian perception grammar (CUE and STRUCT constraints) to account not only for native reading and perception, but also for loanwords from English that show a simultaneous influence of orthography and perception (Sect. 4.3).

The presented reading grammar is thus not restricted to loanword adaptation, and neither is it restricted to Italian. It can be easily applied to other languages with an alphabetic script: Sect. 5.1 below very briefly illustrates how double consonantal letters are treated in native German and in German loanwords from Italian.

Section 5.2 then moves on to show how word recognition works for words that were read via a reading grammar, and how a reading grammar can simply be reversed to account for the process of writing. Section 5.3 discusses alternatives to a mapping from written onto a phonological surface form, and why and when we need to assume such alternative mappings.

5.1 Reading grammars in other languages: German double consonantal letters

The model of a reading grammar introduced in this article is not restricted to Italian but can be applied to all languages with an alphabetic script (but see additional assumptions for irregular scripts such as English discussed in Sect. 5.3 below). In this section, we briefly illustrate how German employs double consonantal letters to indicate the shortness and usually also laxness of the preceding vowel, and how this also holds for non-native words. For our illustration, we use the phonological and orthographic minimal pair in (18), and focus on the contrast in length of the stressed vowel. Note that the German vowel phoneme pair /a/ - /a:/ is one of two that contrast in length, only, cf. Wiese (1996:11). We follow Wiese (1996:35–37) in assuming intervocalic consonants preceded by a stressed short/lax vowel are ambisyllabic in German (see Ramers 1992 for full discussion), represented in this section by a dot under the consonant.

(18) *German minimal pair illustrating vowel length contrast*

- a) /.'ʁaɪ̯.tə./ *Ratte* 'rat'
- b) /.'ʁa:̯.tə./ *Rate* 'rate'

Relevant for our formalization is the fact that German does not allow geminates apart from so-called fake geminates, i.e. two identical consonants that span a prosodic word boundary (Wiese 1996:36). This is captured here with the STRUCT constraint $*\text{GEM}_\omega$ ‘assign a violation mark to every geminate that is not spanning a prosodic word boundary.’ And while phonotactically sequences of a long vowel followed by several consonants are fine, even within one syllable, an interpretation of the orthographic sequence $\langle\alpha\beta_i\beta_i\rangle$ as containing a long/tense vowel is not allowed; see (18a) and monosyllabic words like *matt* /mat/ ‘faint’. In such cases, the writing with two identical consonantal letters encodes a short preceding vowel phoneme. We formalize this with the ORTH constraint $\langle\alpha(\beta_i\beta_i)\rangle/-\text{long}/$. This constraint together with the STRUCT constraint $*\text{GEM}_\omega$ is sufficient to account for the correct reading of (18a), see perception tableau (19). (The following perception tableaux do not formalize the assignment of stress or syllable boundaries.)

(19) *German reading of native <Ratte>*

<Ratte>	$*\text{GEM}_\omega$	$\langle\alpha(\beta_i\beta_i)\rangle/-\text{long}/$	<t>/t/
/.'ʁa:.tə./		*!	*
\mathcal{P} /.'ʁaʦə./			*
/.'ʁat.tə./	*!		

For the correct reading of (18b), we need an additional ORTH constraint to ensure the mapping of $\langle\alpha(\beta\alpha)\rangle$ onto a long/tense first vowel. For this, we employ the constraint $\langle\alpha(\beta\alpha)\rangle/+\text{long}/$, cf. the last column of tableau (20).¹⁵

(20) *German reading of native <Rate>*

<Rate>	$*\text{GEM}_\omega$	$\langle\alpha(\beta_i\beta_i)\rangle/-\text{long}/$	<t>/t/	$\langle\alpha(\beta\alpha)\rangle/+\text{long}/$
\mathcal{P} /.'ʁa:.tə./				
/.'ʁaʦə./				*!
/.'ʁat.tə./	*!			*

For loanwords from languages with geminates that are written with two identical consonantal letters, such as e.g. the Italian words *latte* (*macchiato*), *pizza*, or *ricotta*, the same German reading grammar predicts the correct adapted form, cf. tableau (21).

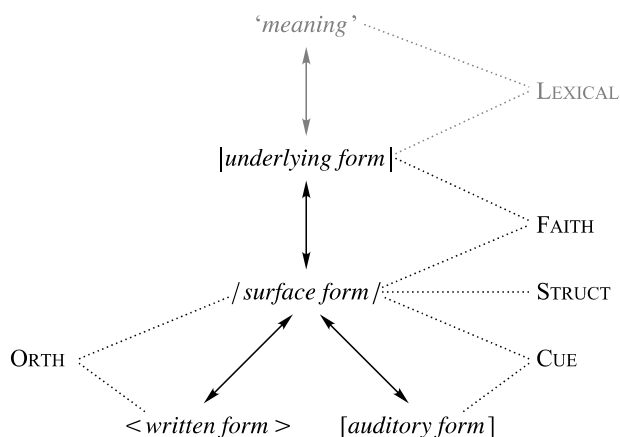
(21) *German reading of non-native <latte>*

<latte>	$*\text{GEM}_\omega$	$\langle\alpha(\beta_i\beta_i)\rangle/-\text{long}/$	<t>/t/	$\langle\alpha(\beta\alpha)\rangle/+\text{long}/$
/.'la:.tə./		*!	*	
\mathcal{P} /.'laʦə./			*	
/.'lat.tə./	*!			

For a full account of geminates in the phonology and orthography of German, see Hamann (submitted).

¹⁵A more general constraint like $\langle\alpha(\beta)\rangle/+\text{long}/$ seems very low ranked in German, because word-final $\langle\alpha\beta\rangle$ sequences are often mapped onto /-long/, see e.g. *hat* /hat/ ‘had’, while long/tense vowels in final syllables usually are specially encoded in writing with two identical vowel letters, e.g. *Saal* /za:l/ ‘hall’, or an additional letter <h>, e.g. *Pfahl* /pfa:l/ ‘stake’.

Fig. 5 The BiPhon model (Boersma 2007) with a meaning level in grey (Apoussidou 2007) and the proposed mapping from written onto phonological surface form (and vice versa), i.e. the sub-lexical route



5.2 The processes of word recognition and writing

Up to now, we considered the reading and perception process both of native and non-native inputs. Boersma's perception grammar that we employed is part of the larger grammar model of *Bidirectional Phonetics and Phonology* (BiPhon; Boersma 2007, 2011), where not only the perception but also the recognition process is modelled, i.e. the activation of a lexical form on the basis of a surface phonological representation. This lexical activation is also necessary as an accompanying step for our proposed mappings from grapheme onto SF. In Sect. 4.1, we mentioned already that the surface forms /.'fat.to./, /.'fa:.to./ and /tʃit.'ta./ we gained from the Italian reading grammar are assumed to be lexically retrieved as [fat:o] 'done', [fato] 'fate' and [tʃit:'a] 'city', respectively. This additional mapping is depicted in Fig. 5 (where the surface form can result either from the orthographic form, the auditory form, or both). The mapping from SF to underlying form (UF) is guided by FAITH constraints (Boersma 2011:37–39),¹⁶ and that from UF to meaning by LEXICAL constraints (Apoussidou 2007; for further distinctions on the meaning level, see Boersma 2011). We refrain from formalizing those upper mappings for the current examples.

In Fig. 5, the arrows connecting all forms point both ways, indicating that the mappings are bidirectional. This bidirectionality is a main principle of the BiPhon model, where not only speech perception and recognition, but also the reverse processes of phonological production and phonetic implementation are modelled. This is done with the same constraints used in the other processing direction: LEXICAL and FAITH constraints map meaning and underlying form onto the phonological surface form in phonological production, where the output is restricted by STRUCT constraints, and CUE constraints are responsible for the mapping from surface onto an auditory form in phonetic implementation.¹⁷

Applying this principle of bidirectional mapping to our reading grammar, we thus also have a *writing grammar* by simply using the ORTH constraints in the opposite

¹⁶Anti-FAITH constraints (Boersma and Hamann 2009:41) possibly also play a role.

¹⁷A further mapping of this auditory form onto an articulatory form is assumed in the process of phonetic implementation. The constraints necessary for its formalization are described in e.g. Boersma (2011).

direction. A constraint as e.g. $\langle t \rangle / t /$ is then interpreted as “assign a violation mark if the phonological surface form $/t/$ is not written as grapheme $\langle t \rangle$.” Tableau (22) is an example of a writing tableau: It has a phonological surface form as input and written forms as output candidates.

(22) *Italian writing of /'fat.to/*

/'fat.to./	PENULT	/.μμ./	$\langle \beta_i \beta_i \rangle / C: /$	$\ast \langle \alpha \rangle / V: /$
\emptyset <fatto>				
<fato>			*!	

The constraints given in this tableau are the same with the same ranking as in reading tableau (8), where we formalized the reading of the orthographic word <fatto>. The first two constraints in (22), PENULT and $/.μμ./$, are STRUCT constraints. They play no role in the writing direction, as they refer to the SF, i.e. the input. In the evaluation of writing, only ORTH constraints are relevant. In tableau (22), the first candidate violates none of the given ORTH constraints and is therefore the winner. The second candidate violates the ORTH constraint $\langle \beta_i \beta_i \rangle / C: /$ because the input SF has a long consonant $/t.t/$ that is not mapped onto a double consonantal letter in the output, the written form.

5.3 Reading as mapping onto a surface form or a higher-level representation?

Many languages with an alphabetic script have inconsistent mappings between graphemes and phonemes, as e.g. in English and French, where the same sound can be written in different ways (heterographs, as e.g. English <here> and <hear>), and the same written form can be pronounced differently (heteronyms, e.g. English <tear> for $[teə]$ and $[tɪə]$). For cases of heterographs and heteronyms, a reading and writing grammar as proposed up to now is insufficient. In order to be able to write, the writer needs to know which of the possible written forms is the one representing the underlying form with the intended meaning. And in order to be able to read, the reader needs to know which of the possible underlying forms and meanings is the one associated with the given written form. To be able to account for such cases, we need an additional way of accessing lexical entries in the reading process. We follow the cognitive *dual-route model* of reading (henceforth: DR model; e.g. Coltheart et al. 1993, 2001) in assuming there are two possible mappings of the written form. One is the *sub-lexical* route, where graphemes are transformed into a so-called ‘sub-lexical phonological form.’ This form corresponds to the SF in our proposal, and we illustrated already how this mapping is performed by ORTH and STRUCT constraints. The second route is the *lexical route*, also known as *direct access*,¹⁸ where the written form is directly mapped onto a lexical entry in the shape of a whole-word phonology together with its meaning. The lexical route in the DR model is explicitly said to employ visual word recognition but no graphemic parsing (Coltheart et al. 1993:597). This means that the reader is mapping the written word in its entirety onto a stored UF word form and a connected meaning, and that phonology plays no role in this

¹⁸ *Direct access* is opposed to *mediated access* via phonological representations; see e.g. Bradshaw (1975) for a review.

holistic access. Such a mapping therefore can also account for logographic writing systems.

A question that arises in this context (and was asked by one reviewer) is whether the mapping we propose in the present article for the sub-lexical route, i.e. from written form onto SF, is not better replaced by a mapping from written onto underlying form (UF). Instead of the ORTH constraints proposed in this article, we would need very similar constraints mapping graphemes onto phonemes in the UF, e.g. <t>|t|: ‘Assign a violation mark to every letter <t> that is not mapped onto a |t| in the UF.’ An argument in favor of such a UF mapping is that most languages with an alphabetic script do not distinguish allophones in their writing system, referring to phonemes instead. One of the few counterexamples can be found in Dutch, where the devoiced allophones of voiced sibilants are encoded in the writing, e.g. <laars – laarzen> ‘boot sg. – pl.’ and <wijf – wijven> ‘woman sg. – pl.’

In the Italian data presented here, we encountered one case of allophony, namely the two allophones /V/ and /V:/ of stressed vowels, which are not distinguished orthographically. For those, we need to assume that Italian readers map the same written form onto two different surface allophones, and subsequently onto one common underlying phoneme. As the above-mentioned reviewer correctly pointed out, it seems more straightforward to assume readers map a vowel grapheme directly onto its corresponding UF. However, we would not be able to restrict this mapping by the same phonotactic constraints that apply in the process of speech perception and phonological production (formalized as STRUCT constraints in the BiPhon model). We would therefore lose exactly what makes our proposal attractive, namely the non-reduplication of such phonological knowledge. Instead, we would have to postulate an additional set of restrictions, this time on the UF. Such *morpheme structure constraints* (MSC) were proposed in rule-based generative phonology (see e.g. Halle 1959; Stanley 1967), but are in conflict with OT’s concept of *Richness of the Base* because they pose restrictions on the input (e.g. McCarthy 1998). Further points of criticism against MSC are that they at least partly reduplicate phonotactic constraints (STRUCT), and seem to capture statistical tendencies rather than absolute constraints (see e.g. Booij 2011 for a full discussion). For these reasons we do not consider a grapheme-UF mapping a valid alternative to our proposal of the sub-lexical route in reading.

Returning to the distinction between lexical and sub-lexical routes, the present model is in agreement with the DR model that nonce words can only be read via the sub-lexical route, as they have no lexical entry. For the same reason, initial orthographic adaptations of loanwords can only occur via the sub-lexical route, as these new words initially lack a lexical form. In the process of reading and thus adaptation, a lexical entry consisting of an underlying phonological form and a meaning is created. In principle it would be possible for a borrower to create a new lexical entry from a written form and a meaning deducted from the context without passing the grapheme-to-phoneme mapping, thus by lexical route. In that case, neither phonotactic restrictions nor phonological mapping constraints would guide this lexical entry, so there would be no restriction whatsoever on the UF.

When reading real words in alphabetic scripts, the lexical and the sub-lexical route are assumed here to be in competition with each other (easily modelled within

OT). This competition is obvious for languages with deep orthographies, for instance English, where the irregularly written <sew> [səʊ], accessible via the lexical route only, competes with a sub-lexical mapping of the grapheme <ew> onto /u:/ as in *few*, *dew*, etc. Competition between the lexical and the sub-lexical route is also commonly found in languages with shallow orthographies (see e.g. Katz and Frost 2001; Grainger et al. 2012).

6 Previous linguistic accounts of orthographic borrowings in Italian and of reading and writing in general

In this section, we compare the present account of singleton/geminate borrowings in Italian to earlier proposals on orthographic borrowings in Italian (Sect. 6.1), and to previous formalizations of the reading and writing process within a grammar theory (Sect. 6.2).

6.1 Earlier accounts of the (non-)gemination of consonants in Italian loanwords

In this subsection, we discuss the studies by Rando (1970), Repetti (1993) and Morandini (2007), as they all looked at borrowings of English intervocalic consonants into Italian, and their observations partly complement and partly challenge the present account. The studies by Passino (2008, 2013) and Repetti (2009, 2012) are not discussed as they predominantly deal with word-final consonants in Italian loans and the question whether these should be treated as singletons or geminates. Though this is an interesting question, it falls outside the scope of the present paper.

Rando (1970) discusses English words that have been introduced into Italian “in both written and oral form” (132), i.e. via orthography and via perception, resulting in two different Italian loanforms (forming cases of *loan doublets* as defined by Smith 2006), cf. the examples in (23) (based on Rando 1970:132; with relevant consonants in boldface).¹⁹

(23)	English word	Borrowed orthographically	Borrowed perceptually
a)	<i>tunnel</i>	/.'tun.nel./	/.'ta:.nel./
	<i>shopping</i>	/.'ʃop.pin./	/.'ʃo:.piŋ./
	<i>tennis</i>	/.'tɛn.nis./	/.'tɛ:.nis./
b)	<i>scooter</i>	/.'sku:.ter./	/.'sku:.tə./
	<i>boomerang</i>	/bo.me.'raŋg./	/.'bu:.me.raŋg./
	<i>reporter</i>	/re.'pɔr.ter./	/re.'pɔ:.tə./

The perceptual borrowings in Rando’s study (last column in (23)) always have singleton consonants, and their stressed vowels are closer to the English pronunciation than those in the orthographic borrowings (second column), as we also observed in our data set. Rando (132) elaborates that the orthographic borrowings he collected are in

¹⁹The transcriptions of the examples given in this section are not those employed in the original sources but our careful transference into the transcription system that we employ throughout the present article (IPA with additional phonological assumptions as elaborated in Sect. 2).

line with the Italian spelling: those written with two consonants are borrowed with a geminate, cf. the examples given in (23a), as opposed to those words with a singleton in (23b). Unfortunately the examples in (23b) all involve a long vowel before the consonant in the English original form. This long vowel (and its written representation with two vowel graphemes in two of the three cases) provides an alternative explanation for their borrowing with a short following consonant. Rando does not comment on loans written with two different intervocalic consonants, though his set of words borrowed with singletons includes *reporter*, which is rendered with an /rt/ sequence in the orthographic borrowing. This is in line with the Italian grapheme-phoneme mappings we propose, as the written sequence <rt> is rendered as /r.t/ via the two Italian ORTH constraints <r>/r/ and <t>/t/, and the resulting bisyllabic sequence does not violate any phonotactic restrictions.

The difference between the orthographic and the perceptual borrowing strategy, Rando remarks, and its resulting variation in borrowing form is sometimes even reflected in two possible writings in Italian, e.g. *pullover/pulover* or *bluff/bluf* (130). For such cases, Rando observes a tendency to preserve the original written form, as it is more prestigious (134). In our data set, far less variation occurred, though we did find considerable variation for the word *pullover* (see Appendix). Rando did not discuss simultaneous orthographic and perceptual adaptations that we covered in Sect. 4.3.

Repetti's (1993) study looks at Italian loanwords and also at Canadian Italian forms, though we focus here on the former to ensure comparability with the data collected in the present article. In her account of why some loans have an open syllable in stressed, penultimate position, and others a closed one, Repetti refers to orthography:

If the written form has a single consonant following the stressed vowel, the vowel is pronounced long in accordance with the rules of the Italian writing system. If, instead, the stressed vowel is followed by two written consonants, that vowel is pronounced short. (Repetti 1993:192)

Though the focus in her analysis is on vowel length, she makes the same generalization with respect to writing as Rando (1970). And as in Rando's dataset, Repetti's examples illustrating short consonants (after long vowels) all involve originally long vowels (or diphthongs) in English, cf. (24a), and therefore could also be accounted for by perceptual vowel length instead of orthography.

(24) *Loanwords from Repetti (1993:192) illustrating stressed penultimate vowels*

- a) *computer* /.'kom'pju:.ter./
- shaker* /.'ʃeɪ.ker./
- slogan* /.'zlo:.gan./
- smoking* /.'zmɔ:.kiŋ./
- b) *flipper* /.'flip.per./
- tunnel* /.'tun.nel./
- budget* /.'bad.dʒet./
- check-in* /.'tʃɛk.kin./

The examples in (24b) illustrate the correlation between a borrowed geminate and a written form with two consonants. As we can see, Repetti's generalization for those

is not restricted to two identical consonantal letters (similar to Rando's), and the last two words in (24b) seem to be counterexamples to the restriction we proposed in the present article. For the first word, *budget*, both Morandini (2007:21) and Zingarelli et al. (2015) also give /.'bad.dʒet./, and our native speakers vary in their judgments (see Appendix). The intervocalic grapheme sequence <dg> can be read as a sequence of coda /d/ followed by an onset /dʒ/, which would lead to a pronunciation that is indistinguishable from a geminate /d:ʒ/. If we assume such a biphonemic form, then this word does not form an exception to an adaptation via the reading grammar.²⁰

The second alleged counterexample, *check-in*, differs from the loans that we discussed in that it consists of a nominalized phrasal verb, with a morpheme boundary right after the relevant consonant in the English form. This example and similar nominalizations of phrasal verbs that we found in Zingarelli et al. (2015) are given in (25).

(25) *Borrowings of nominalized phrasal verbs (from Zingarelli et al. 2015)*

<i>check (-in/up)</i>	/tʃɛk./	1966
<i>blackout</i>	/blek.'aut./	1949
<i>knockout</i>	/nok.'kaut./	1911
<i>pick-up</i>	/pi.'kap./	1931
<i>backup</i>	/be.'kap./	1988

Our native speakers produced all of these words with singletons, apart from *check-in*, which one older speaker produced as /tʃɛk.kin./ (see Appendix). According to Morandini (2007:22), *backup* and *check-in* are borrowed with geminates. The variation observed in the borrowings of these nominalized phrasal verbs seems to depend on whether Italian native speakers analyze them as consisting of one or two syntactic words. If analyzed as consisting of two syntactic words, where the first is stressed and ends in a consonant and the second begins with a vowel, the sequence is likely to undergo *backwards raddoppiamento* (Chierchia 1986), as is e.g. the case in *tram elettrico* /'tram.me.'let.tri.ko./ 'electric tramway' or *gas asfissiante* /'gas.sas.fis.'sjan.te./ 'asphyxiating gas' (Repetti 1993:189). Variation between borrowed forms with singleton or geminate in such cases could thus be assigned to a difference in syntactic interpretation, which explains the exceptional status of /tʃɛk.kin./ in (24b).²¹

The study by Morandini (2007) on Italian loans from English focuses on stress assignment and consonant clusters. He observes that Italian speakers borrow what he calls "graphic geminates" (27), i.e. two identical consonantal letters, as geminates, independent of their realization in the donor language. Morandini's examples illustrating borrowings with geminates are also part of our data set in (3). His examples with singletons are given in (26):

²⁰We would like to thank one of the three anonymous reviewers for pointing this out to us.

²¹Morandini (2007:22) reports two forms for the loan *weekend* /wi.'kɛnd./~/.wik.'kɛnd./ (Zingarelli et al. 2015 gives only /wi.'kɛnd./), which might be due to the same reason, namely application of backwards raddoppiamento only if the word is interpreted as consisting of two separate syntactic words by the borrower.

- (26) *Orthographic borrowings with singletons from Morandini (2007:27)*
- | | |
|----------------|--------------|
| <i>pony</i> | /.'pɔ:.ni./ |
| <i>beauty</i> | /.'bju:.ti./ |
| <i>meeting</i> | /.'mi:.tiŋ./ |
| <i>hacker</i> | /.'a:.ker./ |

The first three examples in (26) are borrowings of English forms with a long vowel before the consonant in question, and therefore allow a perceptual explanation of the adaptation (see our criticism of Rando 1970 and Repetti 1993 above). The fourth example, however, involves a short vowel and furthermore a sequence of two differing consonant letters, and thus supports our observation that such sequences are not borrowed as geminate.

Morandini's data involve three further borrowings (21) that are of interest to the present account. The first is the word *access*, which despite its perceptual form with two differing intervocalic consonants [ks] has been borrowed orthographically with a geminate as /.'at.tʃɛs./, and thus confirms our proposal that two identical consonantal letters are interpreted as geminates (if there is no perceptual input to correct for this).²² The other two borrowings involve cases of two identical consonant letters followed by a further consonant, given in (27).

- (27)
- | | | |
|------------------|--------------------|-----------------------------|
| <i>dribbling</i> | /.'drib.bliŋ./ | |
| <i>bulldozer</i> | /.'bul.'dɔd.dzer./ | (see also Repetti 1993:190) |

While the orthography would predict a geminate in both cases, only the first word is in accordance with this prediction. Here, the following consonant is an /l/. Laterals can follow geminate plosives in Italian; therefore the orthographic adaptation does not violate any Italian phonotactic constraint. In the case of *bulldozer*, on the other hand, the following consonant is a plosive, which is not allowed in post-geminate position (as discussed in Sect. 2). In addition, the single grapheme <z> in *bulldozer* is realized as geminate /dʒz/ due to the non-existence of a singleton alveolar affricate in intervocalic position. In this word we thus encounter two further cases of phonotactic restrictions (*/.ld/ and */VdʒV/) overriding the orthographic mapping, similar to the cases of *fashion* and *puzzle* discussed and formalized in Sect. 4.2.

As mentioned already in the introduction, none of the studies on English loanwords in Italian known to us provide a formalization of the role of orthography in a grammar model, a point we will focus on in the next section.

6.2 Previous linguistic models of reading and writing

The earliest linguistic accounts of orthography that employ an explicit formalization can be found in the tradition of rule-based Generative Grammar, and employ formal rewriting rules similar to those used for the description of phonological processes. Bierwisch (1972), for example, provides context-sensitive orthographic rewriting rules that take phonological feature matrices as input and generate written forms from them. These early rule-based approaches to orthography are concerned with the production, i.e. the writing process, only, and consider the written form to be solely

²²Note that the second author of this article does not recognize *access* as a loanword.

derivable from phonology. This idea was taken to the extreme by Chomsky and Halle (1968), who propose that in English the underlying form very often is identical to the written form.

A newer and somewhat different rule-based approach to orthography is provided by Neef (2012), who formalizes the reading process (which he calls “recoding”). Neef distinguishes between individual correspondence rules (e.g. $\langle m \rangle \rightarrow [m]$) and constraints that capture general properties of the writing system. An example of such a general constraint for German is, e.g. “in a sequence of identical letters, all non-initial ones may be recoded as zero” (211).²³ For the written input $\langle mm \rangle$, this constraint provides an alternative output $[m]$ in addition to $[mm]$. All outputs generated by correspondence rules and constraints are checked for their phonological well-formedness via a phonological filter, which discards ungrammatical outputs (e.g. $[mm]$ at the end of a syllable). Though Neef’s principle that phonology “controls” the reading process is in line with the present account, his use of rules to create outputs make additional machinery in the form of constraints and a phonological filter necessary, and also requires a derivation with at least two stages. In the present OT formalization, on the other hand, mapping constraints (ORTH) and restricting constraints (STRUCT) simply interact in the choice of the best candidate. Furthermore, though Neef writes that his recoding system can also be used for the writing process, it is not clear how this is performed with the provided correspondence rules, especially those that include context, as they cannot be simply reversed, while we showed in Sect. 5.2 how the proposed ORTH constraints work bidirectionally, i.e. are interpretable both in the reading and writing direction.

A full-fledged OT account of writing is provided by Wiese (2004), who proposes sound-letter correspondence constraints that map underlying phonological input onto written form. The evaluation is performed with the help of correspondence constraints such as MAX, DEP and IDENT, which are usually employed in OT to evaluate the mapping between UF and SF (Prince and Smolensky 1993 [2004]), and in Wiese’s account incorrectly imply a possible identity between written and phonological form. In the present account, the different nature of the two forms is made explicit by employing constraints that arbitrarily map one onto the other. Wiese furthermore distinguishes between predictable and unpredictable features of an orthographic system, where the former are dealt with by constraints, whereas the latter are stored in the lexicon. Such a distinction is, however, difficult to make in languages with irregular orthographies. Though Wiese’s account is restricted to writing, he points out that the “bidirectional nature of correspondence relations allows for constraints looking into both directions” (316). Wiese’s account is extended in the study by Song and Wiese (2010), who propose that the input to the OT orthographic evaluation can be either the UF (e.g. in Korean) or the SF (what they call “lexical outputs” 91; e.g. in German).

Baroni (2013) in his account of alternative writings for existing English words (e.g. $\langle \text{tonite} \rangle$ for *tonight*) proposes bidirectional constraints that map the SF onto written forms and vice versa. An example is his constraint $\langle VCe \rangle \leftrightarrow \text{TENSEV}$, which stands for “ $\langle V \rangle$ in $\langle VCe \rangle$ sequences is bidirectionally mapped onto a tense vowel” (31).

²³This constraint captures the fact that German orthography uses two identical consonantal letters to indicate the shortness/laxness of the preceding vowel, which we encoded with the ORTH constraint $\langle \alpha(\beta_i \beta_i) \rangle / -\text{long} /$ in Sect. 5.1.

These constraints are then used to model writing only, where phonotactic restrictions do not play any role.

In terms of formalizing orthography especially for loanword adaptation, only the study by Dong (2012) deals with this topic. Dong, referring to the BiPhon model, proposes OT constraints of the type “<x>_{ORTH} should not be mapped to /y/_{SF}” (48) for the correlation between Pinyin written forms and Mandarin Chinese surface forms. However, Dong neither employs such constraints in her OT formalization of Mandarin borrowings, nor discusses their possible interactions with other (such as *STRUCT*) constraints.

In sum, none of the previous proposals make a principled distinction between orthography and phonology, or show how the two interact in the reading (or writing) process.

7 Conclusion

In this article we proposed that the borrowing of English intervocalic consonants after short/lax vowels into Italian is influenced by orthography in such a way that only consonants that are orthographically represented with two identical letters are borrowed as geminates, whereas those represented with a single letter or a sequence of two different consonantal letters are incorporated as singletons. We formalized the orthographic borrowing in an OT framework with the help of *ORTH* constraints, responsible for grapheme-to-phoneme mappings, and argued that this mapping is influenced by phonological, i.e. structural restrictions (*STRUCT* constraints) on Italian. Together, the language-specific ranking of *ORTH* and *STRUCT* constraints form the native Italian reading grammar, which was shown to be able to handle the reading of native words. The application of this native reading grammar to non-native written forms from English was shown to account for the attested borrowed forms. Compared to earlier proposals, our study is trailblazing, providing a first formal account of adaptations via orthography, and making the role of phonotactic restrictions in the reading and writing grammar explicit without reduplicating these restrictions in the orthographic constraints. Such adaptation does not require any loanword-specific devices but rather makes use of the reading grammar necessary for reading and writing of native Italian. Furthermore, we showed with an example from German that the proposed model is not language-specific but rather applicable to all languages that employ an alphabetic script. Though not illustrated here, the proposed reading grammar can be easily extended to languages that use syllabaries as writing systems, such as Japanese kana, where *ORTH* constraints map syllabograms onto syllables that are, in turn, restricted by *STRUCT* constraints. For logographic scripts, such as Chinese characters (hanzi), we proposed direct mappings (the so-called lexical route) from written form onto pairs of underlying form and meaning. We leave elaboration of these reading grammars and corresponding writing grammars for future work.

In fact, our study uncovers several topics still to be dealt with. First, as mentioned at the beginning of Sect. 4, English orthography can only influence Italian borrowings because both languages employ a Roman alphabetic script. English uses several graphemes that do not, or once did not, occur in Italian orthography such as <k>,

<th> and <ck> (and, conversely, Italian has several graphemes that do not occur in English). A native reading grammar can only be applied to graphemes that are used in the native writing system. At the same time, new graphemes *can* be introduced via simultaneous perceptual and orthographical borrowings, as occurred with <k> in Italian, which has been substituting <ch> even in native words (Hall 1958), e.g. *kilometro* 'kilometer' as alternative to *chilometro*. How the introduction of a new grapheme proceeds is another topic we leave for future work.

Furthermore, we did not address the possible knowledge Italian native speakers might have of English orthography through L2 acquisition of English. The quality of the stressed vowels in a word such as *buffer* /.'baf.fer./ that we explained as a perceptual effect, for instance, could also be explained by assuming that the borrower had some knowledge of English grapheme-to-phoneme mappings for the vowel.²⁴ Since the borrowing of the two intervocalic consonantal letters as a geminate is clearly caused by Italian orthography, we would deal with an incomplete L2 reading grammar for English (e.g. a high-ranked constraint <u>/a/) complemented by the native Italian reading grammar, which could be formalized in our proposed reading grammar as an interaction of two language-specific rankings of universal ORTH constraints (or, in an emergentist approach, as an interaction of language-specific Italian and English ORTH constraints).

A further topic of interest we only briefly touched upon is the possible difference between orthographic and perceptual borrowings. In her study of loan doublets in Japanese, Smith (2006) found that orthographic borrowings lead to more epenthesis whereas perceptual borrowings were likelier to result in segmental deletion. In our case, orthography also led to epenthesis of phonological material, namely of a second mora for intervocalic consonants represented with two identical consonantal letters, despite these consonants being monomoraic in the phonological structure of the source language English. We did not find any cases of deletion for perceptual borrowings. In our data, there seemed to be a different asymmetry between the two borrowing strategies. We observed a bigger influence of auditory information on the borrowing of vowels, whereas the borrowing of consonants seemed more influenced by writing. This might be due to the larger perceptual salience of vowel cues compared to consonantal cues. Connected to this topic is the question of whether perception, orthography or both guide Italian speakers in their borrowing of two adjacent vowel letters, such as in *account* /au/, *hockey* /ei/ or *hippie* /i/ (all from our dataset in (3) and (4)). Diphthongs in stressed syllables seem to favor perceptual adaptation, such as in *account*, but also *mouse* /au/, *meeting* /i:/ and *leader* /i:/, while diphthongs in unstressed syllables favor orthographic adaptation, such as in *austerity* /au/ and *hockey* /ei/ (Zingarelli et al. 2015). This adaptation of <VV> sequences and its formalization are also left for analysis in future research.

Acknowledgements We would like to thank our participants for their participation. Furthermore, we thank Paul Boersma, Robert Cloutier, Edoardo Cavarani, Jonathan Weinand, and the audience at the *Manchester Linguistics and English Language Research Seminar*, the *Workshop on the occasion of Marko*

²⁴To reiterate, the grapheme-phoneme correspondences in English are quite irregular. The grapheme <u> can represent, for example, /ʌ/ /ʊ/, /u:/, /ɜ:/, /ɛ/ or /ə/, though it corresponds to /ʌ/ more often than to any of the others (see the corpus study by Berndt et al. 1987:8).

Simonović's defence and the *23rd Manchester Phonology Meeting* and the three reviewers for helpful comments.

The first author is responsible for the analysis and writing up of the results, the second author for the data collection and for an initial analysis with orthographic constraints that contained phonological information, see Colombo (2014).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix: Responses by native speakers

If no alternative form is given, then the form provided by Zingarelli et al. (2015) was used by all six speakers.

Word	Zingarelli et al. (2015)	Alternative forms used by 6 tested native speakers
<i>banner</i>	/.'ban.ner./	
<i>hobby</i>	/.'ɒb.bi./	
<i>horror</i>	/.'ɔr.ror./	
<i>jolly</i>	/.'dʒɔl.li./	
<i>hippie</i>	/.'ip.pi./	
<i>shopping</i>	/.'ʃɒp.pin./	
<i>thriller</i>	/.'tril.ler./	
<i>rally</i>	/.'rel.li./	
<i>splatter</i>	/.'splat.ter./	
<i>newsletter</i>	/.'njuz.'let.ter./	
<i>jogging</i>	/.'dʒɔg.gin./	
<i>happening</i>	/.'ep.pe.nin./	~/'ap.pe.nin./
<i>baby sitter</i>	/.'be.bi.'sit.ter./	
<i>buffer</i>	/.'baf.fer./	
<i>no comment</i>	/.'no.'kɒm.ment./	
<i>attachment</i>	/.'at.'tʌtʃ.mɛnt./	
<i>commando</i>	/.'kom.'man.do./	
<i>account</i>	/.'ak.'kaunt./	Stress-shifted /.'ak.kaunt./ by one younger speaker
<i>pullover</i>	/.'pul.'lɔ:.ver./	Singleton /.'pu.'lɔ:.ver./ ~ /.'pu.'lɔ:.ver./ by one younger speaker and one older speaker; a second younger and a second older speaker found singleton pronunciation an acceptable alternative
<i>hockey</i>	/.'ɔ:.kei./	
<i>hacker</i>	/.'a:.ker./	Geminate form /.'ak.ker./ was accepted as alternative by one older speaker
<i>editor</i>	/.'e:.di.tor./	Form with tensed vowel /.'e:.di.tor./ was used by all young and one older speaker
<i>glamour</i>	/.'gle:.mur./	
<i>cameraman</i>	/.'ka:.me.ra.men./	
<i>monitor</i>	/.'mɔ:.ni.tor./	
<i>fashion</i>	/.'fɛʃ.fon./	
<i>puzzle</i>	/.'pa:.zel./	One older native speaker also accepts /.'pud.dzel./ and /.'pud.dzle./; the other two older speakers and two young speakers accepted /.'pa:.zol./ in addition to /.'pa:.zel./
<i>mobbing</i>	/.'mɔ:.bin./	All six speakers used /.'mɒb.bin./

References

- Apoussidou, Diana. 2007. *The learnability of metrical stress*. PhD diss., University of Amsterdam.
- Bafille, Laura. 1999. Antepenultimate stress in Italian and some related dialects: Metrical and prosodic aspects. *Rivista di Linguistica* 11: 201–229.
- Baroni, Antonio. 2013. Eye dialect and casual speech spelling: Orthographic variation in OT. *Writing Systems Research* 5: 24–53.
- Bassetti, Bene, Paola Escudero, and Rachel Hayes-Harb. 2015. Second language phonology at the interface between acoustic and orthographic input. *Applied Psycholinguistics* 36: 1–6.
- Berndt, Rita, James Reggia, and Charlotte Mitchum. 1987. Empirically derived probabilities for grapheme-to-phoneme correspondences in English. *Behaviour Research Methods, Instruments and Computers* 19: 1–9.
- Bertinetto, Pierre Marco, and Michele Loporcaro. 2005. The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of the International Phonetic Association* 35: 131–151.
- Bierwisch, Manfred. 1972. Schriftstruktur und Phonologie. *Probleme und Ergebnisse der Psychologie* 43: 21–44.
- Boersma, Paul. 1997. How we learn variation, optionality, and probability. *Institute of Phonetic Sciences of the University of Amsterdam (IFA)* 21: 43–58.
- Boersma, Paul. 2007. Some listener-oriented accounts of h-aspiré in French. *Lingua* 117: 1989–2054.
- Boersma, Paul. 2011. A programme for bidirectional phonology and phonetics and their acquisition and evolution. In *Bidirectional Optimality Theory*, eds. Anton Benz and Jason Mattausch, 33–72. Amsterdam: Benjamins.
- Boersma, Paul, and Silke Hamann. 2009. Loanword adaptation as first-language phonological perception. In *Loan phonology*, eds. Andrea Calabrese and W. Leo Wetzels, 11–58. Amsterdam: Benjamins.
- Boersma, Paul, and Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32: 45–86.
- Booij, Geert. 2011. Morpheme structure constraints. In *The Blackwell Companion to Phonology. Volume 4: Interfaces*, eds. Marc van Oostendorp, Colin Ewen, Elizabeth Hume, and Keren Rice, 2049–2070. Oxford: Blackwell.
- Bradshaw, John L. 1975. Three interrelated problems in reading: A review. *Memory and Cognition* 3: 123–134.
- Cattinelli, Isabella, N. Alberto Borghese, Marcello Galluci, and Eraldo Paulesu. 2013. Reading the reading brain: A new meta-analysis of functional imaging data on reading. *Journal of Neurolinguistics* 26: 214–238.
- Chierchia, Gennaro. 1986. Length, syllabification and the phonological cycle in Italian. *Journal of Italian Linguistics* 8: 5–33.
- Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.
- Colombo, Iliaria. 2014. Italian loanword adaptation in OT: How orthographic representation affects perception. Ms., University of Amsterdam.
- Coltheart, Max, Brent Curtis, Paul Atkins, and Michael Haller. 1993. Models of reading aloud: Dual-route and parallel-distributed processing approaches. *Psychological Review* 4: 589–608.
- Coltheart, Max, Kathleen Rastle, Conrad Perry, Robyn Langdon, and Johannes Ziegler. 2001. DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review* 108: 204–256.
- Daland, Robert, Mira Oh, and Seyjeong Kim. 2015. When in doubt, read the instructions: Orthographic effects in the adaptation and online perception of English vowels in Korean. *Lingua* 159: 70–92.
- Detey, Sylvain, and Jean-Luc Nespoulous. 2008. Can orthography influence second language syllabic segmentation? Japanese epenthetic vowels and French consonantal clusters. *Lingua* 118(1): 66–81.
- D'Imperio, Mariapaola, and Sam Rosenthal. 1999. Phonetics and phonology of main stress in Italian. *Phonology* 16(1): 1–28.
- Dong, Xiaoli. 2012. *What borrowing buys us: A study of Mandarin Chinese loanword phonology*. PhD diss., University of Utrecht.
- Escudero, Paola, Rachel Hayes-Harb, and Holger Mitterer. 2008. Novel second-language words and asymmetric lexical access. *Journal of Phonetics* 36: 345–360.
- Escudero, Paola, and Karin Wanrooij. 2010. The effect of L1 orthography on non-native vowel perception. *Language and Speech* 53(3): 343–365.
- Esposito, Anna, and Maria Gabriella Di Benedetto. 1999. Acoustical and perceptual study of gemination in Italian stops. *Journal of the Acoustical Society of America* 106(4): 2051–2062.

- Flege, James Emil, and Ian MacKay. 2004. Perceiving vowels in a second language. *Studies in Second Language Acquisition* 26: 1–34.
- Flege, James Emil, Ian MacKay, and Dian Meador. 1999. Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America* 106(5): 2973–2987.
- Friesner, Michael L. 2009. The adaptation of Romanian loanwords from Turkish and French. In *Loan phonology*, eds. Andrea Calabrese and W. Leo Wetzels, 115–130. Amsterdam: Benjamins.
- Grainger, Jonathan, Bernard L  t  , Daisy Bertand, St  phane Dufau, and Johannes C. Ziegler. 2012. Evidence for multiple routes in learning to read. *Cognition* 123: 280–292.
- Hall, Robert A. Jr. 1944. Italian phonemes and orthography. *Italica* 21(2): 72–82.
- Hall, Robert A. Jr. 1958. Kappa pubblicit  rio. *Lingua Nostra* 19: 129.
- Halle, Morris. 1959. *The sound pattern of Russian: A linguistic and acoustical investigation*. The Hague: Mouton.
- Hamann, Silke. Submitted. One phonotactic restriction for speaking, listening and reading: The case of the no geminate constraint in German. Ms., University of Amsterdam.
- Hamann, Silke, and David W. L. Li. 2016. Adaptation of English onset clusters across time in Hong Kong Cantonese: The role of the perception grammar. *Linguistics in Amsterdam* 9: 56–76.
- Kang, Yoonjung. 2009. English /z/ in 1930s Korean. In *2nd International Conference on East Asian Linguistics*, eds. David Potter and Dennis Storoshenko. Vol. 2 of *Simon Fraser University Working Papers in Linguistics*.
- Katz, Leonard, and Laurie Feldman. 1983. Relation between pronunciation and recognition of printed words in deep and shallow orthographies. *Journal of Experimental Psychology* 9: 157–166.
- Katz, Leonard, and Stephen Frost. 2001. Phonology constrains the internal orthographic representation. *Reading and Writing* 14: 297–332.
- Kr  mer, Martin. 2009. *The phonology of Italian*. Oxford: Oxford University Press.
- Liberman, Isabelle, Alvin Liberman, Ignatius Mattingly, and Donald Shankweiler. 1980. Orthography and the beginning reader. In *Orthography, reading, and dyslexia*, eds. James R. Kavanagh and Richard L. Venezky, 67–84. Baltimore: University Park Press.
- Liberman, Isabelle, and Donald Shankweiler. 1985. Phonology and the problems of learning to read and write. *Remedial and Special Education* 6: 8–17.
- Loporcaro, Michele. 1990. On the analysis of geminates in Standard Italian and Italian dialects. In *Natural phonology: The state of the art. Papers from the Bern Workshop on Natural Phonology*, eds. Bernhard Hurch and Richard Rhodes, 149–174. Berlin: de Gruyter.
- Marotta, Giovanna. 1988. The Italian diphthongs and the autosegmental framework. *Certamen Phonologicum* 8: 389–420.
- McCarthy, John J. 1998. Morpheme structure constraints and paradigm occultation. In *Annual meeting of the Chicago Linguistics Society (CLS)*, Vol. 32, 123–150.
- Miao, Ruiqin. 2005. Loanword adaptation in Mandarin Chinese: Perceptual, phonological and sociolinguistic factors. PhD diss., State University of New York, Stony Brook.
- Morandini, Diego. 2007. *The phonology of loanwords into Italian*, MA thesis, University College London.
- Neef, Martin. 2012. Graphematics as part of a modular theory of phonographic writing systems. *Writing Systems Research* 4: 214–228.
- Passino, Diana. 2008. *Aspects of consonantal lengthening in Italian*. Padova: Unipress.
- Passino, Diana. 2013. A unified account of consonant gemination in external sandhi in Italian: *Raddoppiamento Sintattico* and related phenomena. *The Linguistic Review* 30: 313–346.
- Pickett, Emily, Sheila Blumstein, and Martha Burton. 1999. Effects of speaking rate on the singleton/geminate consonant contrast in Italian. *Phonetica* 56: 135–157.
- Porter, Stacey. 2010. *Orthographic influence on the perception and production of Spanish loans in English*, MA thesis, University of California, Irvine.
- Prince, Alan, and Paul Smolensky. 1993 [2004]. *Optimality theory: Constraint interaction in generative grammar*. Malden: Blackwell.
- Ramers, Karl Heinz. 1992. Ambisyllabische Konsonanten im Deutschen. In *Silbenphonologie des Deutschen*, eds. Peter Eisenberg, Karl H. Ramers, and Heinz Vater, 246–283. T  bingen: Narr.
- Ramus, Franck, and Gayaneh Szenkovits. 2008. What phonological deficit? *Quarterly Journal of Experimental Psychology* 61(1): 129–141.
- Rando, Gaetano. 1970. The assimilation of English loan-words in Italian. *Italica* 47: 129–142.
- Repetti, Lori. 1993. The integration of foreign loans in the phonology of Italian. *Italica* 70(2): 182–196.
- Repetti, Lori. 2009. Gemination in English loans in American varieties of Italian. In *Loan phonology*, eds. Andrea Calabrese and W. Leo Wetzels, 225–239. Amsterdam: Benjamins.

- Repetti, Lori. 2012. Consonant-final loanwords and epenthetic vowels in Italian. *Catalan Journal of Linguistics* 11: 167–188.
- Rogers, Derek, and Luciana d'Arcangeli. 2004. Italian. *Journal of the International Phonetic Association* 34: 117–121.
- Saltarelli, Mario. 1983. The mora unit in Italian phonology. *Folia Linguistica* 17: 7–24.
- Saltarelli, Mario. 1984. Italian syllable structure. In *Estudis Gramaticals: Working Papers in Linguistics*, Vol. 1, 279–295. Barcelona: Universitat Autònoma de Barcelona.
- Smith, Jennifer. 2006. Loan phonology is not all perception: Evidence from Japanese loan doublets. In *Japanese/Korean Linguistics 14*, eds. Timothy Vance and Kimberly Jones, 63–74. Stanford: CSLI.
- Song, Hye Jeong, and Richard Wiese. 2010. Resistance to complexity interacting with visual shape—German and Korean orthography. *Writing Systems Research* 2: 87–103.
- Stanley, Richard. 1967. Redundancy rules in phonology. *Language* 43: 393–436.
- Vendelin, Inga, and Sharon Peperkamp. 2006. The influence of orthography on loanword adaptations. *Lingua* 116(7): 996–1007.
- Vogel, Irene. 1982. *La sillaba come unità fonologica*. Bologna: Zanichelli.
- Wiese, Richard. 1996. *The phonology of German*. Oxford: Oxford University Press.
- Wiese, Richard. 2004. How to optimize orthography. *Written Language and Literacy* 7: 305–331.
- Zingarelli, Nicola, Mario Cannella, and Beata Lazzarini. 2015. *Lo zingarelli 2016: Vocabolario della lingua italiana*. Bologna: Zanichelli.